Erkenn (2012) 76:1–22
DOI 10.1007/s10670-011-9329-4

ORIGINAL ARTICLE

Causation Without Influence

Tomasz Bigaj

Received: 4 December 2008 / Accepted: 1 October 2011 / Published online: 19 October 2011

Abstract David Lewis's latest theory of causation defines the causal link in terms of the relation of influence between events. It turns out, however, that one event's influencing another is neither a necessary nor sufficient condition for its being a cause of that event. In the article one particular case of causality without influence is presented and developed. This case not only serves as a counterexample to Lewis's influence theory, but also threatens earlier counterfactual analyses of causation by admitting a particularly troublesome type of preemption. The conclusion of the article is that Lewis's influence method of solving the preemption problem fails, and that we need a new and fresh approach to the cases of redundant causation if we want to hold on to the counterfactual analysis of causation.

1 Introduction

David Lewis's counterfactual analysis of causation started off inconspicuously as a modest alternative to the dominating regularity approach, but it soon grew into the mainstay of modern philosophy of causation, overshadowing its regularity rival. What may come as a surprise (but only a mild one, given the nature of philosophical investigations) is the fact that Lewis's approach owes its enormous popularity not only to its successes, which are unquestionable, but first and foremost to the number of problems it creates. Since its conception in the seventies, the counterfactual analysis of causation has been subjected to a tremendous amount of critical analysis from various authors, even including Lewis himself. Under this pressure the counterfactual theory has undergone substantial transformations, modifications, and amendments, and these changes in turn have elicited even more sophisticated counterexamples and objections. Although not all of Lewis's off-hand attempts to

T. Bigaj (✉)
Institute of Philosophy, Warsaw University, Krakowskie Przedmiescie 3, 00-047 Warsaw, Poland
e-mail: t.f.bigaj@uw.edu.pl

overcome numerous obstacles have withstood the test of time, some of them have given rise to better developed alternatives, of which the most recent is the influence theory of causation that appeared in print in 2000, 1 year prior to his untimely death. It has to be noted that Lewis's last word on the issue of causality did not escape the fate of his earlier analyses—the criticism leveled against the influence theory of causation is no less intense than that directed against its predecessors, and we can only regret that it won't be met with Lewis's sharp and witty responses.

The purpose of this paper is primarily to discuss one particular counterexample to Lewis's latest theory of causation. The counterexample that will be subsequently analyzed in the paper has some interesting features that set it apart from typical examples known in the literature.[1] First of all, the same single case with only slight modifications can serve as a falsifier of virtually every earlier version of Lewis's counterfactual theory, including the variants rejected by Lewis himself. The required modifications involve preempted backup causes, and it turns out that this new type of preemption is particularly resistant to the standard available methods of dealing with redundant causation. This fact dashes any hope of quickly saving the counterfactual analysis by adopting either a simple or qualified 'disjunctive approach', according to which each definition (or modification thereof) given by Lewis can serve as a partial analysis of causation, applicable in certain circumstances where others may fail.[2] Moreover, the counterexample selected for analysis in this paper is astonishingly simple—one may even call it embarrassingly simplistic. It does not involve any potentially objectionable elements such as omissions, preventers, magical spells, science-fiction contraptions, quantum non-local interactions, etc., that pervade the current literature on the subject. Thanks to this feature the counterexample cannot be lightly dismissed, as the intuitive judgment on which it rests is virtually unassailable.

But even more significantly, the example selected as a trouble-maker to Lewis's theory turns out to belong to an important subcategory of causation that is quite common in the physical world, and yet to my knowledge has not received the required share of attention. I base this judgment on the fact that this particular subcategory of causal connections between events seems to falsify not only Lewis's analysis of causation, but also some of the approaches that promise to take physical causality seriously, such as Salmon's or Dowe's analyses in terms of the transfer of a conserved quantity. Finally, the supplied example has the potential to shed new light on the distinction between causal factors and causal conditions.

But first let us give a whirlwind presentation of the main developments in the turbulent history of the counterfactual approach to causation.[3] The entire adventure

[1] An anonymous referee has pointed out that an example virtually identical to the one I develop in Sect. 3 has been mentioned in the book (Schwarz 2009, p. 144). I do not wish to make any claim of authorship of this example, but I would like to make it clear that at the time of writing the first version of this paper in 2008 I did not know about Schwarz's important book. I am grateful to the referee for correcting my oversight.

[2] It has to be noted that Lewis himself was skeptical about the theoretical advantages of the disjunctive approach to causality (Lewis 2004, p. 76).

[3] Two fundamental texts that laid the foundations for the modern counterfactual theory of causation are (Lewis 1973, 1986). (Collins et al. 2004) gives an excellent overview of the history of the counterfactual approach.

began with the simple analysis based on the notion of counterfactual dependence. Stripped down to its bare essentials the simple counterfactual analysis stipulates that an actual event c is a cause of another actual event e distinct from c in case it is true that were c not to occur, e would not occur either. However, such a definition is unable to ensure the transitivity of the causal relation due to the fact that counterfactual conditionals do not satisfy the law of transitivity. For that reason Lewis decided to define the causal relation as the ancestral of the relation of counterfactual dependence. This amendment helped Lewis deal with the first serious challenge to his theory, namely the case of early preemption.

Early preemption involves two (potential) causes c_1 and c_2 of an event e such that the occurrence of one of them (let's say c_1) not only causes e to occur, but also cuts off the causal chain that leads from c_2 to the same effect e. As a result, c_2 becomes a preempted potential cause, whereas c_1 remains an actual cause of e. However, it is not true that had c_1 not occurred, e would not have happened, for in such a case e would have been produced by the backup cause c_2. Lewis solves this problem neatly by singling out an intermediate event d in the chain of events connecting c_1 with e, such that d occurs *after* c_2 has been preempted. In such a case we can argue that d is counterfactually dependent on c_1, while e is counterfactually dependent on d, hence the amended definition is satisfied.[4] Note that to secure the truth of the counterfactual $\sim O(d) \,\Box\!\!\rightarrow\, \sim O(e)$ we have to assume that so-called backtracking counterfactuals are not allowed, in order to forestall the argument that if d had not happened, c_1 must not have happened, and therefore c_2 would have happened. Given this anti-backtracking assumption, Lewis's analysis seems to easily clear the first hurdle. But there is more to come.

A more threatening challenge, which ultimately led to the downfall of the ancestral-of-the-counterfactual-dependence analysis, is the case of late preemption. In this case c_1 does not preempt c_2 until e is produced, which makes it too late to use the trick with the intermediate event d. The following standard example can illustrate this situation: two children, Suzy and Billy, are throwing rocks at a bottle, but Suzy throws her rock a moment earlier than Billy, and therefore it is her rock and not Billy's that shatters the bottle. However, until Suzy's rock shatters the bottle Billy's flying rock threatens to produce the same effect. Hence there is not a moment before the shattering of the bottle when we could truthfully say "Had Suzy's rock not been on its way, the bottle would not have shattered".

Lewis considered several possible strategies to deal with this obstinate case. For some time his preferred solution was based on the notion of quasi-dependence. Lewis noticed that there are intrinsic copies of the entire causal chain leading from Suzy's throw to the bottle's shattering for which the counterfactual dependence of the effect on the cause is preserved—these are namely situations in which Billy is out of the picture. Due to the existence of such an intrinsic copy with the counterfactual dependence, we can call the relation between the bottle's actual shattering and Suzy's actual throw "quasi-dependence", and Lewis stipulates that this relation should be sufficient for causation to occur. However, later he

[4] Of course this solution is not available to the prominent group of philosophers who deny that causation is transitive. See e.g. (Hall 2000) and a response given in (Lewis 2004).

abandoned this solution, apparently as a result of insurmountable problems it entailed.

Critics have pointed out that it is possible to find cases in which an intrinsic copy of a causal chain with counterfactual dependence is not itself causal, thus refuting Lewis's claim on which the purported solution rests.[5] To that we may add that if the notion of quasi-dependence is meant to serve as part of the definition of the causal relation, then this clearly leads to circularity, as quasi-dependence in turn is based on the notion of causal chains. And it looks like the reference to a *causal* chain, and not to any other chain (for instance a spatiotemporally contiguous one), in the definition of quasi-dependence is unavoidable, otherwise we would have a problem of how to exclude the preempted cause from the process whose intrinsic copy is supposed to display the counterfactual dependence between the cause and the effect.

Another possible way of dealing with the late preemption case, briefly examined by Lewis but quickly dismissed, is the solution based on the notion of fragile events. Lewis observes that strictly speaking the shattering of the bottle by Suzy's rock is not identical in every respect with the shattering caused by Billy, as either rock was thrown from a slightly different direction, possibly at a different speed, angle, etc. Hence if we adopted very strict criteria of identity for the actual event of the bottle's shattering, we could claim that without Suzy's throw *this particular* event of shattering would not have happened, although a very similar one would have taken place instead. This solution presupposes that events are very *fragile* entities, as a seemingly insignificant change in a given event's properties could result in its 'destruction' and the subsequent creation of a numerically different event. But Lewis points out that the proposed medicine is worse than the disease, as it leads to an enormous proliferation of *spurious* causes. Fragile events are counterfactually dependent on numerous earlier occurrences, the majority of which are not intuitively considered genuine causes. For instance, in the case of Suzy's throw, a slight gust of wind could minimally affect the trajectory of the rock and thus the pattern of the shattering, and that fact leads to the desperate conclusion that if the wind had not blown, the shattering (as it happened) would not have occurred. Faced with such a consequence, Lewis unsurprisingly abandons the strategy based on the

[5] McDermott (1995) gives such an example involving double prevention: stopping a mad American President from initiating a nuclear conflict with the Soviet Union is a cause of an ordinary Joe Blow's eating his breakfast the next morning, as it prevents Soviet missiles from preventing the breakfast. But McDermott observes that an intrinsic copy of all relevant events in America does not constitute a causal chain in the world in which Russia is uninhabited or in which all Russian nuclear missiles are made of papier-mâché. However, it is not exactly clear to me why the actual conditions in Russia are to be excluded from the entire causal chain leading from the restraining of the crazed President to Joe Blow's breakfast. I believe that *intrinsicness* is an altogether different concept than *locality* or *contiguity*. An intrinsic copy of a process has to share with it all non-relational properties of the original process, but the process itself may consist of spatiotemporally separated elements (conditions). To put it differently, the state of alertness of the military in Russia may be *extraneous* to what is happening in America, but is not an *extrinsic* matter with respect to the connection between the restraining and the breakfast. Other counterexamples to the quasi-dependence theory are presented by Lewis himself in (2004, pp. 83–85). I believe all of them can raise legitimate questions, but I will not pursue this matter any further.

notion of fragility.[6] The only reason I mention this ill-conceived solution to the late preemption worry is that it ultimately gave rise to a better conception based on the notion of influence. Interestingly, however, some critics maintain that similar examples of spurious causality may also affect the latter proposal. We will return to this shortly.

Even the fragility solution is unable to solve yet another conundrum affecting the counterfactual analysis of causation, known as 'trumping preemption'. In this case we are supposed to imagine a major and a sergeant shouting the same order to a soldier. The soldier carries out the order, but given the hierarchy of military ranks it is natural to assume that it was the major's and not the sergeant's order that ultimately caused the soldier's action. But even assuming that the soldier's action is a fragile event it is difficult to argue that if the major had not shouted the command, the executing of the order would be numerically different from the actual one.[7] Lewis ultimately admitted that his early counterfactual analysis is unable to cope with this problem, and therefore he decided to try a new approach, using the notion of influence.

2 The Influence Theory of Causation

Lewis's earlier analysis of causality focuses exclusively on one type of counterfactual dependence between events: the so-called whether–whether dependence. But there are other types of dependence, not limited to the crude all-or-nothing dependence of occurrences. For instance, we can speak about when–when dependence, when-whether dependence, and the broader how–how dependence. Generally, not only the existence of one event may depend on the existence of another event, but also the time and manner of the occurrence of one of them can counterfactually depend on the time and manner of the occurrence of the other.

[6] A noteworthy attempt to revive the fragility solution has been made by Coady (2004). He defends a position according to which in the real world there exist both fragile and robust versions of common events. The expression "the shattering of the bottle" is ambiguous, as it can refer either to a fragile or a robust event. When we say that Suzy's throw was a cause of the shattering, we interpret the shattering as a fragile event and thus we accept the counterfactual dependence between it and the throw. On the other hand, the robust shattering is counterfactually dependent on both throws taken jointly, but it doesn't admit spurious causation. Coady considers and repels various objections to his solution, including the charge of an unnecessary multiplication of distinct events. While unquestionably interesting, Coady's solution has some obvious shortcomings. It is far from clear that the statement giving the causal explanation of the shattering is indeed considered ambiguous in natural language. Coady claims that there are contexts in which it is intuitively acceptable to deny that Suzy's throw alone caused the shattering, but I simply don't see this. It seems to me that there is only one natural answer to the question "What caused the shattering?", where the shattering is understood in the most general and unspecific way, and the answer is unambiguously "Suzy's throw did that".

[7] Coady (2004) suggests a solution based on his distinction between fragile and robust versions of the major's order. If we interpret the major's order as a fragile event, then the closest possible world in which this particular order does not exist is arguably not a world in which the major does not give any order, but a world in which the major gives a slightly different order. In such a world the soldier will obey the major's new order, and thus the counterfactual dependence between the soldier's action and the major's order is restored.

These loose thoughts can be given a more rigorous explication with the help of Lewis's notion of influence (Lewis 2000, 2004).

Lewis stipulates that an event c influences an event e if and only if there is a substantial range R_c of different not-too-distant alterations of c and a range R_e of alterations of e, such that each alteration of e within R_e is counterfactually dependent on some alteration of c from R_c. And by an alteration of an event x Lewis understands either a very fragile version of x or a very fragile alternative event that is similar to x but numerically different. The idea of introducing the notion of alteration comes from the unsuccessful attempt to solve the problem of late preemption by making the effect a fragile event. Lewis does not want to commit himself to the interpretation of actual events which makes them excessively fragile, but instead introduces an array of extra non-actual and fragile events that 'surround' the actual one. The fact that those extra events may or may not be seen as numerically identical with the actual one reflects Lewis's intention to leave the issue of fragility of actual events open. His new approach to causality is supposed to be neutral with respect to that problem.

Finally, Lewis stipulates in the standard fashion that the causal relation is the ancestral of the relation of influence. Now we can see how the influence theory deals with the problem of late preemption. When we consider Suzy's throw, we can introduce a set of possible events that are similar to the actual throw, but differ with respect to some parameter, such as the exact time of the throw, the initial speed, direction, the rotation of the rock, etc. And it turns out that to each such event there corresponds a fragile alteration of the shattering of the bottle such that if the initial event had not occurred, this particular alteration would not have occurred either. To use Schaffer's happy phrase, if we 'wiggle' Suzy's throw a bit, the shattering will wiggle accordingly. On the other hand, there is no such dependence between the alterations of Billy's rock and the alterations of the shattering. The time and manner of Billy's throw does not influence the time and manner of the shattering, at least not in a significant way (we have to disregard effects such as the minute gravitational impact of Billy's rock on the bottle). In a similar way we can dissolve the problem of trumping preemption, arguing that it is the major's order and not the sergeant's that influences the soldier's action.

However, one may point out that our analysis of Suzy and Billy's target practice was a bit biased toward the required solution. To see that let us assume that Billy's actual throw took place 1 s later than Suzy's. Now, it can be claimed that in fact the definition of influence is satisfied by the pair (Billy's throw, the bottle's shattering) when we take the following set as the range of alterations of the first of the two events: the set of Billy's alternative throws that occurred more than a second prior to the actual throw, and that differed slightly from the actual one with respect to some other parameters. It is clear that this set is mapped onto appropriate alterations of the shattering, for in this case it is Billy's rock and not Suzy's that reaches the bottle first. Given the vagueness of Lewis's phrase "not-too-distant alterations", a reasonably strong case can be made for the conclusion that Billy's throw influences the shattering of the bottle, thus undermining the appeal of the new influence theory of causation.

In order to avoid similar problems, Igal Kvart (2001) has proposed to make Lewis's concept of influence a bit more precise and accurate with the help of some topological notions. The general idea is to prescribe that if there is a pattern of influence between a range of alterations R_c that constitute a *connected neighborhood* of the actual event c and a range of alterations R_e, then if we take a narrower connected neighborhood R'_c included in R_c and still containing c, this pattern of influence should not degenerate into a mapping that funnels all alterations from R'_c onto one actual event e. In other words, we should be able to make the alterations of the cause c arbitrarily small, and still get non-trivial alterations of the effect. This condition is arguably not satisfied in the case of Billy's throw, because if we take a sub-range of alterations of his throw whose time of occurrence deviates less than 1 s from the actual event, then the pattern of influence will become degenerate (in that case it is still Suzy's rock that reaches the bottle first).

But even this correction to Lewis's account does not eliminate all potential troubles. Critics have been quick to point out other cases that apparently satisfy Lewis's definition and yet intuitively do not exemplify the causal relationship. Dowe (2000a, b) for instance invokes examples of *hasteners* and *delayers*, i.e. events that make another event occur earlier or later than it would have occurred otherwise, but do not qualify as causes. An example of a delayer may be a rain that causes a subsequent forest fire to occur later than it would have occurred without the rain. Schaffer (2001) describes a case of an electrocution using a device that has two controllers: one is a single on–off button, and the other is a complicated switchboard that affects all parameters of the electrocution (its time, duration, the amount of the flowing current, etc.) but does not turn on the device. Intuitively, only the operation of the main button counts as a cause of the electrocution, but the relation of influence clearly holds between the setting of the switchboard and the event of electrocution.

One possible response to these counterexamples is to concede, contrary to our off-hand snap judgement, that they constitute legitimate cases of causation after all. Lewis has already made one step toward broadening his notion of cause by accepting causal conditions as genuine causes. If we agree that the presence of oxygen is one of the causes of the forest fire, then what is wrong with saying that even an earlier rain could have causally contributed to it? Of course rains are generally not conducive to fires, as opposed to oxygen, but in one particular case a rain can actually help a fire occur at a given time (rather than at another). But still, we may wonder if this laxity in admitting diverse causes of a given event does not go too far. Even Lewis was reluctant to accept a gust of wind as a cause of the bottle's shattering, or eating a big meal before taking a lethal dose of poison as a cause of death. That is why he voted to reject the fragility solution to the problem of late preemption. But now it seems that the same problem comes back again in his own approach. Can't we say that the gust of wind influences the shattering? After all, if we consider alterations of this particular gust, then clearly they will be mapped onto some minute alterations of the way the bottle shattered.

Still, the situation is far from being clear. The counterexamples presented so far can certainly shake our confidence in the new theory of causation, but they fall short of dealing a fatal blow. By adopting a certain threshold of what counts as a

legitimate alteration of the effect we could try to eliminate most egregious cases while leaving borderline cases as 'spoils to the victor'. But the arsenal of troublesome cases is not exhausted yet. We haven't discussed yet apparent situations in which we can have a causal relation without influence. Schaffer for instance claims that in his electrocution device example pressing the button does not stand in Lewis's relation of influence to the electrocution. True, he admits that the alterations of the time at which the device is turned on are associated with the alterations of the time of the electrocution, but he insists that this when–when dependence is not sufficient to satisfy Lewis's definition of influence, as not a single parameter of the electrocution other than its time depends on any parameters of flipping the switch. But to forestall possible complaints he even modifies the entire set-up in such a way that the change of the time of the switch's flipping does not produce the change of the timing of the electrocution. Although the situation becomes more complicated because of that modification, the main goal seems to be achieved—we have a case of a cause-and-effect link without a pattern of influence.

Kvart conceived of a similar example involving a high-tech modification of Suzy and Billy's target practice. Suppose that Suzy's rock, rather than shattering the bottle directly, sets off a complicated electronic detection device that is connected to a small explosive attached to the bottle. Again, it can be claimed that no parameter of the shattering caused by the explosion other than its time (and even this can be eliminated) depends on any parameters of Suzy's throw, but clearly Suzy's throw is a cause of the shattering. I believe that both Schaffer's and Kvart's examples raise a legitimate question about whether the influence conception of causation is able to account for all cases of causation that we are intuitively inclined to accept.[8] But to settle the issue decisively let us consider the following low-tech and run-of-the-mill example.

3 The Railroad Switch Counterexample

Let us consider one of the simplest imaginable set-ups: a railroad track that splits into two tracks, and a switch regulating the direction of train traffic. Suppose further that one of the two tracks leads to a dead end, and that at 6:00 PM a passenger train is supposed to pass this fork on the way to its destination. However, at 5:00 PM a bad guy creeps in and changes the position of the switch by moving the mechanical lever that operates the switch, so that the coming train is now directed onto the dead end track. As bad luck would have it, no one notices the change, and the train rolls full speed onto the wrong track, crashing at the end of it. If any intuitive causal judgments are to be taken seriously by philosophers of causality, it is definitely the

[8] Another counterexample to Lewis's theory of influence has been put forward by Tooley (2003, p. 409). Instead of a particular example he considers a general situation involving three actual events C, D and E. By assumption there is no relation of influence between C and E, but we assume that in the absence of C, D does influence E, meaning that an appropriate range of alterations of D (D_1, D_2, …) is mapped onto respective alterations of E (E_1, E_2, …). However, when C and any of the alterations of D occur jointly, the result is always the same effect E. Tooley claims that in this case it is most natural to accept that C is a cause of E that preempts D, and yet C does not satisfy Lewis's condition.

claim that moving the switch was a cause of the crash (although for sure other causal factors, including the train's motion, had to contribute to this regrettable outcome as well). And yet a moment's reflection can reveal that there is no possible way we could claim that flipping the lever influences the crash, according to Lewis's definition of the notion of influence. You are free to imagine all sorts of alterations of the actual way the bad guy moved the lever that fall short of making this move unsuccessful—moving the lever at 5:05, moving it at 4:55, moving it with the left hand, kicking it, moving it slowly and deliberately, moving it lightning fast, etc.—and not a single one of these alterations would make the slightest difference in the way the train crashed. Try as you might to change the way the lever was moved, and you'll never be able to get any change in the crash. There is simply no dependence between the crash and the flipping of the switch other than the pure whether–whether dependence.

But maybe not all hope is yet lost for Lewis's analysis of causality in terms of influence. Remember that the causal relation is not defined directly as the relation of influence, but rather more flexibly as its ancestral. So couldn't we find an intermediate event between the moving of the lever and the train's wreck such that this event is influenced by the former and influences the latter? Let us try.

There is definitely an event that is connected to the movement of the lever via Lewis's relation of influence—it is namely the lateral movement of the switch rails (these are two linked rails that lie between the diverging rails and can be moved into one of the two positions to determine the direction in which the train goes). Surely the time and manner of the pulling of the lever determines the time and manner of the switch rails' motion. But at this point the chain of influence stops dead. Again, varying the time and manner of the motion of the switch rails does not in the least affect the time and manner of the crash, nor the time and manner of any event that involves the train passing through the switch.

You may point out that there are some special alterations of the actual movement of the switch rails that are mapped into a dramatic alteration of the crash. Consider, for instance, those alterations that leave the switch rails in the middle and inaccurate position between two diverging tracks (supposing that the mechanism of the switch allows for such a flawed execution of the change), and therefore cause the train to derail on passing the switch. True, but this fact does not ensure that the corrected definition of the relation of influence is satisfied, for at least two reasons. Firstly, this mapping is still a degenerate one, as it maps several 'incomplete' movements of the switch rail onto one and the same result: the early derailment. Secondly, and more importantly, taking smaller and smaller alterations of the actual event we will end up with the switch rail being put in the correct position, which leads to the same crash as in reality. But, according to Kvart's amendment, the pattern of influence should be preserved even under arbitrarily small departures from the actual cause. For that reason I don't see any reasonable way to make Lewis's definition work in this case.

The railroad switch example goes in the same direction as the two counterexamples given by Kvart and Schaffer, but makes the case against Lewis's theory of influence even more compelling, due to its simplicity, the strength of the intuitions on which its relies, and the fact that the when–when dependence is excluded

naturally at the outset, without any need of artificial add-ons.[9] Further advantages of this example will be revealed in due course, but at this point let us already observe that it can also cast doubts on one potential solution to the problem of how to define causation, suggested by Schaffer at the end of his critical article (2001, p. 18).

Without elaborating, Schaffer makes a comment that the most promising way of approaching causation is in terms of the relation of *effluence* that, roughly, involves a process connection between the cause and the effect. Without knowing the details of this proposal it is difficult to judge, but it seems to me that at least one natural interpretation of Schaffer's notion of effluence is in terms of the transfer of some dynamical quantities (such as energy or momentum). Schaffer's electrocution case can be satisfactorily analyzed using the notion of effluence, because it is natural to assume that there is a continuous physical process that connects the pressing of the button and the electrocuting device.[10] But this solution does not work in the railroad case. The moving of the lever is connected neither by Lewis's influence nor Schaffer's effluence to the resulting crash. No dynamical parameters that characterize the shifting of the lever are relevant to the dynamical parameters of the crash. On the other hand, the relationship of effluence certainly holds between the coming train and its crash. Hence it seems that Schaffer's suggested approach is capable of delivering one correct verdict that the oncoming train was a cause of the crash, but has difficulties with singling out another cause of it—namely the changing of the switch. What distinguishes these two causes of the crash—the movement of the train and the changing of the switch's position—is that the latter creates a necessary background condition for the crash, while the former initiates the dynamical process that due to the 'right' conditions is able to lead to the final outcome. But between the movement of the lever and the train's crash there is a long period of 'dormancy' in which no dynamical process takes place (other than simply objects persisting in time) until the train arrives and is directed onto the dead end track.

But it may be pointed out that my purported counterexample can be easily dismissed, as it clearly displays the whether–whether dependence, and the existence of this sort of dependence is arguably sufficient for the relation of influence to hold. Lewis himself states explicitly that "the simplest sort of whether–whether dependence, with only two different alterations of E [i.e. the effect], still qualifies as one sort of pattern of influence" (2004, pp. 91–92), and later he adds that "absences can be among the unactualized alterations of a cause or effect that figure in a pattern of influence" (*ibid.*, p. 100).[11] However, these claims are difficult to

[9] While Schaffer's counterexample seems to be equally simple and compelling, it nevertheless requires some artificial additions to eliminate the whether-whether and when-when dependence of the electrocution on the pressing of the button. No such addition is necessary in the example considered above. The when-when dependence does not exist between the flipping of the switch and the train wreck, whereas the whether-whether dependence can be easily eliminated by a back-up cause, as explained below.

[10] On second thought, Schaffer's claim can be questioned too. In a sense the pressing of the button acts in a similar way to the flipping of the railroad switch: it closes an electric circuit, allowing the current to flow from the source of electricity to the electrocuting device. The pressing itself is a mechanical process whose physical parameters do not stand in any direct relation to the physical parameters of the electrocution.

[11] I am indebted to an anonymous referee for pressing this point.

accept given his notion of influence. Lewis's remarks don't seem to square with both the letter and spirit of his theory of influence. According to his earlier stipulation, an alteration of an event is just a modified version of this event, or a numerically distinct event which is nevertheless similar to it (the choice between these two options depends on whether we consider the initial event to be fragile or not). But how can the absence of an event be its version, or a closely resembling counterpart of it? Lewis explains that when we suppose counterfactually that an event *C* does not occur, we should "imagine that *C* is completely and cleanly excised from history, leaving behind no fragment or approximation of itself" (*ibid*., p. 90). This, in my mind, speaks decisively against interpreting absences of events as their alterations, and consequently against including whether–whether dependence in the pattern of influence.[12]

Incidentally, it is worth noticing that Lewis's own example showing how to incorporate the whether–whether dependence into the broader influence pattern differs significantly from the railroad switch case. Lewis notes that sometimes small alterations of the cause are mapped into only two alterations of the effect (i.e. its presence and absence). This type of "extreme funneling" can be illustrated with the help of the following double prevention case: a fighter airplane escorting a bomber shoots down an enemy airplane preventing it from preventing the bomber from destroying the target. Some alterations of the shooting of the interceptor by the escort are clearly mapped into an absence of the effect (if the escort fires and misses the enemy plane by a millimeter, the interceptor will most probably shoot down the bomber and the target will not be reached). But our current case is different. Small alterations of the cause do not get mapped into anything other than the actual train crash. Only a big alteration of the flipping of the switch (i.e. not flipping it at all) can influence the outcome. Even if we granted to Lewis that alterations may contain absences as their special cases, still his definition of the relation of influence contains the phrase "not too distant" which is applied to the alterations of the cause, but not of the effect. So in assessing whether the influence relation holds between two events we are not allowed to consider absences of the purported cause, because they count as too distant alterations.

But of course it is always possible to combine Lewis's two approaches—the earlier counterfactual whether–whether dependence approach and the latest one based on the notion of influence—together in a disjunctive fashion, and thus achieve the needed flexibility. Under such an approach preemption cases are accounted for with the help of the relation of influence, while the cases exhibiting no influence but the whether–whether dependence are still analyzable in terms of the counterfactual dependence. Seemingly we lose nothing by this marriage, and we gain a lot. However, the railroad switch example can be easily modified in order to block the whether–whether dependence with the help of a good old redundant and preempted cause. Will the Lewisians be able to deal with this modification using any of the available strategies other than the influence approach?

[12] A similar point is made in (Schwarz 2009, p. 145).

4 Back to Preemption

Let us consider the following correction to the story about the bad guy flipping the switch. Suppose that the terrorist cell responsible for planning the attack had decided to play it safe and, unbeknownst to the first guy, sent along a backup conspirator in order to observe the action of the first one. The idea is of course that if bad guy number one fails to do the job, the second one is supposed to jump in and move the switch. In this new set-up the whether–whether dependence between the lever's shifting and the train's crash is broken, since had the first guy not moved the switch, the second guy would have done this and the very same crash would have occurred. But this scenario looks very similar to the early preemption case, as the successful execution of the switch's change by the first guy clearly obviates the need of any action from the backup conspirator. Hence this case should be solvable with the help of the ancestral of the relation of counterfactual dependence. But is it?[13]

To fix the ideas, let us discuss two variants of the backup scenario, reflecting two different specific modi operandi of the second conspirator. In the first scenario, which can be called an 'early intervention', the backup guy is supposed to wait until just seconds after 5:00 PM, and if the main executor of the plot is nowhere to be seen, the backup has to act immediately. According to the second possible modus operandi ('late intervention'), the backup plotter has to wait until just before 6:00 (the time of the train's arrival at the switch), and if the switch is still not moved, he is obliged to pull the lever. Let us start with the early intervention case.

In order for Lewis's strategy to work, we have to find an intermediate event between the actual pulling of the lever at 5:00 and the train's wreck, such that this event counterfactually depends on the pulling. But this is not an easy task. The backup plotter is supposed to move the same lever at virtually the same time as the actual executor. Hence whatever actual event following the moving of the lever we consider, it is not true that had the first guy not moved the lever, this event would not have happened. Unless we have grounds to believe that the second plotter would have moved the lever in a significantly different way than his accomplice, we cannot claim that any subsequent actual event depends counterfactually on the actual moving of the lever. Thus Lewis's strategy fails even before it can get off the ground.

The late intervention case fares no better with this respect. True, now we can find plenty of events such that had the lever's shift at 5:00 PM not occurred, they would not have occurred either. These are namely all the events involving the position of the switch rails before 6:00 PM. However, because the backup guy is supposed to act just a moment before the train's arrival, the crash would have happened anyway, so it does not depend counterfactually on any event earlier than 6:00 PM. On the other hand, no event happening after the scheduled time of the backup intervention is counterfactually dependent on the actual action of the first guy. Again, as in the

[13] Of course we could assume from the outset that the backup assassin operates in the late pre-emption mode. For instance, he might have installed a device which, upon the arrival of the train, moves the switch in the required position if it is not already set in this position (see later in the main text). But it is interesting to see that even the early pre-emption version does not admit the usual solution advocated by Lewis in his earlier works.

early intervention scenario, the backup plotter is supposed to move the same lever and thus create the same condition of the switch, which together with the train's arrival leads to the same crash. So, for instance, it is not true that if the first guy had not moved the lever, the train would not have rolled onto the dead end track, and similarly for all subsequent events leading to the disaster.

A quick reflection reveals that nothing significant changes if we adopt a mid-time intervention scenario. No matter what time *t* between 5:00 and 6:00 the backup assassin chooses as the moment of his intervention, had the first guy failed to act at 5:00, it is always true that no event after *t* (including the final effect) is counterfactually dependent on any event before *t* (including the actual cause), and this fact makes it impossible for Lewis's ancestral-of-counterfactual-dependence definition to be satisfied. But why? What is it about this particular case of early preemption that makes it different from the standard example with Suzy and Billy throwing rocks at a bottle?

Recall that an early preemption variant of this story is that Billy, upon seeing Suzy throwing her rock, decides not to throw his rock, but had she failed to throw it, he would have shattered the bottle with his rock. In this case all we have to do is choose a moment at which Billy gave up on his desire to strike the bottle, when Suzy's rock was still on its way to the target. Clearly, at this moment the presence of Suzy's rock in mid-flight is counterfactually dependent on her toss, and the shattering is counterfactually dependent on the rock's being on its way (given no backtracking). But the situation with the railroad switch is not entirely analogous. The reason for this is that the backup guy would initiate the process that goes through the same channel that the actual process that started at 5:00 PM. Every single event belonging to the possible backup process is qualitatively identical with an appropriate event in the actual process. This sets the railroad switch example apart from the Suzy and Billy case, in which each child uses his/her own channel (different rocks and different trajectories) to bring about the same effect.

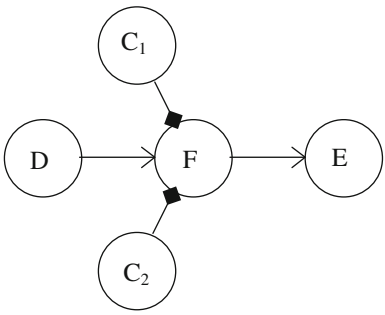
In order to make their situation more in line with the switch example, we would have to imagine a scenario according to which Billy, upon seeing that Suzy decided not to throw her rock, would jump to her, grab her rock from her hands and throw it exactly the same way she actually did. But even in this case any delay in executing Suzy's intended throw would result in a different chain of events leading to the cause, thus allowing for the standard solution to the early preemption case. A nice feature of the switch example is that the causal process goes through the period of 'dormancy' (the switch's being in the new position and waiting for the train to arrive), where it can be assumed that each event in the temporal series is qualitatively indiscernible from the previous one (for instance, the state of the switch's being in the dead-end-track position at 5:25 is qualitatively the same as the state of the switch's being in the dead-end-track position at 5:55).

The solution using the notion of fragility will not work here for the same reason that the influence theory does not work, namely because the crash caused by the backup terrorist would be exactly the same as the crash caused by the first plotter. The only available strategy left is the solution based on the notion of quasi-dependence. Notwithstanding Lewis's and others' doubts regarding this strategy, let us notice that in this particular case it seems to work: after all, it is easy to imagine

an exact intrinsic copy of the entire situation from the initial shift of the lever till the train's crash, without the backup conspirator. But in order to repel this last desperate attack on our counterexample let us introduce a high-tech modification. Rather than a human backup, let us imagine that the terrorists managed to plant a special device that operates as follows: when the coming train reaches its detector, it sends a signal to the switch, checking its position. If the position directed the train away from the dead end track, the device would move the switch, otherwise it would remain idle. Suppose, moreover, that the device sends a unique electric impulse to the switch that no other electronic system can emulate. This signal leaves an 'imprint' on the microstructure of the switch rails. In such a case any intrinsic copy of the actual chain of events will have to posses this imprint, and therefore will have to contain the device that acts as a backup. The counterfactual dependence between the mechanical operation of the switch and the crash cannot be restored. Consequently, there is no quasi-dependence between the cause and the effect. The case seems to be closed.

5 The Neuronal Network Example

Since Lewis's 'Postscripts to "Causation"' (1986) it has become standard to use artificially designed examples involving neuronal links in order to illustrate various possible causal set-ups. This approach relieves us from the complications associated with real-life cases that may obscure our view and draw attention from important matters to irrelevant details. It turns out that it is relatively easy to design a situation involving neuronal links that is analogous to the railroad switch example with redundancy. All we have to do is to introduce a new type of neuron whose role is to prepare normal neurons to fire when stimulated (we can call this preparation 'priming'). On the diagram below the connection between neuron C_1 and F, shown by a line with a diamond at its end, represents such a priming link. When C_1 fires F becomes ready to receive a stimulating signal from D and to fire as a result of this stimulation, but without priming F will not fire even when stimulated by D. Moreover, we assume that another priming signal (such as the one received from neuron C_2) does not significantly change the state of neuron F if it is already primed.



Now we can consider the following sequence of events: at t_1 neuron C_1 fires and primes F, at t_2 another priming neuron C_2 fires, but this does not change the state of

neuron F, and at some later time t_3 neuron D fires a stimulating signal that in turn causes F to fire and stimulate neuron E. Obviously D's firing counts as a cause of E's firing, but so does the firing of C_1, according to our pre-theoretical intuitions, for C_1 creates a necessary condition required for F to be able to pass the signal through to neuron E. The firing of C_2 in turn is a preempted cause of E. The simple counterfactual analysis yields an incorrect verdict regarding C_1, as there is no counterfactual dependence between E and C_1 (if C_1 had not fired, F would have been primed by C_2, and D would have stimulated F to fire). But now we can observe that none of the various strategies devised by Lewis to cope with different cases of preemption will work in this case. The quasi-dependence solution fails in this case, as it can be used to argue for the incorrect conclusion that the firing of C_2 is a cause of the firing of E. We have assumed that the firing of C_2 does not change the state of the already primed neuron F in any significant way. If that is the case, then we can imagine an exact intrinsic copy of the sequence of events leading from the firing of C_2 to the firing of E with neuron C_1 eliminated. In this situation there is obviously a counterfactual dependence between the firing of E and the firing of C_2, hence the condition of quasi-dependence is satisfied for the firings of neurons C_2 and E.[14]

Finally, neither the fragility solution nor the most recent influence theory can deliver the correct analysis of the entire situation. We can reasonably argue that the relation of influence links the firing of D with the firing of E. This is due to the fact that changes of the time of D's firing and (presumably) of its intensity will yield appropriate changes of the time and intensity of E's firing. However, no such link exists between the firing of C_1 and the effect. As long as an alteration of the actual firing of C_1 takes place before t_3 and its intensity does not fall below the threshold required for F to be primed, this alteration will be mapped onto exactly the same firing of neuron E as in actuality.

6 Causes and Conditions

One conceivable strategy to refute the above-mentioned examples and to save Lewis's analysis is to try to apply the distinction between direct causal factors and causal conditions. It may be claimed that the firing of neuron C_1 is not strictly speaking a cause of the firing of E, as it creates only the prerequisite conditions. But the conditions themselves are incapable of bringing about the effect unless a true 'dynamical' cause (in the form of the firing of neuron D) occurs. Regardless of the plausibility of such a defense we have to note that it does not help much in furthering Lewis's views of causation, since his analysis of the simpler case without the backup neuron C_2 clearly implies that the firing of C_1 *is* a cause of the firing of E. Lewis himself insisted that the distinction between causal factors and causal conditions does not carry much theoretical weight, and that it can be relegated

[14] Lewis in (2004, p. 83) describes a similar counterexample to the quasi-dependence solution which is a modification of the standard Suzy and Billy example. However, in order to consider a possible world in which an intrinsic copy of the sequence from Billy's throw to the shattering constitutes a causal chain, Lewis has to invoke laws that allow ordinary objects to "jump" randomly at distances. On the other hand, my example with priming neurons works perfectly well in the worlds with ordinary laws.

entirely to the sphere of pragmatics. But I believe that we could, contra Lewis, make this important distinction less nebulous and pragmatic with the help of some theoretical concepts definable in the semantic framework of counterfactual logic.[15]

In my (Bigaj 2005) I have suggested that what distinguishes background conditions from causes is that although the effect is (typically) counterfactually dependent on both, a background condition is usually *cotenable* with the absence of the effect while a cause is not. This fact can be expressed equivalently in the requirement that for an event c to be a cause and not only a background condition of e, it has to be true that if e had not occurred, c *might* not have occurred. Now I believe that for various reasons this requirement may not be entirely acceptable, but I maintain that a more flexible distinction can still be drawn.[16] We could make the distinction between background conditions and causes a matter of degree rather than a yes or no issue, with the help of a simple possible-world comparison.

Let us take an exemplary case illustrating the distinction in question: the striking of the match versus the presence of oxygen as a cause (or a condition) of the lighting of the match. It is true that in the closest world in which there is no oxygen the match will not light, precisely as in the case of the closest possible world in which there is no striking. But we have good reasons to believe that there is an important difference between the two cases: the worlds in which there is no oxygen in the atmosphere surrounding the match seem to be more remote from the actual world than the worlds in which there is no striking of the match. Thus it looks like a reasonable criterion to accept that of two events c_1 and c_2 such that event e is counterfactually dependent on either of them, c_1 is more a genuine cause and c_2 more a background condition if the closest not-c_1-worlds are closer to the actual world than the closest not-c_2-worlds.

But now it can be observed that the firing of neuron C_1 in our previous example looks more like a genuine cause in comparison with typical background conditions. Take, for instance, the fact that at t_3 neuron F existed. True, if neuron F had not existed at t_3, E would not have fired, as there would have been a gap between neurons D and E. However the existence of neuron F is not seen as a cause of the firing of neuron E, but rather as its background condition. And now it is reasonable to maintain that the world in which neuron F ceases to exist at t_3 is more remote than the world in which neuron C_1 fails to fire. We don't have to devise fancy criteria of similarity between possible worlds, involving the distinction between big and small miracles, in order to argue that this is the case. There are many natural scenarios that involve the failure of C_1 to fire—it may be due to some electrical interference from the surrounding neuronal network, it may be because of a lack of some vital chemical compound, etc. But to account for the non-existence of neuron F we would

[15] Peter Menzies is of the same opinion. He tries to account for the difference between causes and background condition using his concept of 'difference-making in context' (Menzies 2004).

[16] My addition to Lewis's definition of cause would work only under the assumption that backtracking is allowed, for only in such a case we could say truthfully that if there was no effect, its (earlier) cause might not have occurred. I believe that backtracking counterfactuals are perfectly legitimate in such contexts, but I don't have space here to consider this problem in detail. Another problem with my approach is that it does not seem to work well under the assumption of indeternimism. For details and possible remedies see (Bigaj 2005, pp. 607–610).

have either to go far back in time to the moment when the entire neuronal network was developed to make necessary adjustments, or imagine some super-natural way of eliminating the existing neuron. On the other hand, the possible failure of neuron C_1 to fire seems to be on a par with the possible failure of neuron D to fire with respect to their remoteness from the actual world, and therefore both actual firings qualify as genuine causes. Hence the distinction between causes and background conditions cannot help to repel the counterexample to Lewis's analysis.[17]

7 Whither Counterfactual Analysis?

But how serious is the problem we have presented? I believe it is rather serious, due to the fact that examples of causal links of the sort we have presented are quite common in real life. Causation without influence is present in many situations that we encounter in everyday life, and it also occurs commonly in science. To give a few examples: you insert a bullet in an empty magazine of a revolver, and later somebody accidentally shoots himself while cleaning the revolver. You move the electric switch in the on position during a power outage, and later when the power is restored the light bulb flashes. You dissolve a bit of salt in chemically pure H_2O, and later electric current is passed through it, electrocuting somebody. Inserting the bullet, moving the switch, dissolving salt, all qualify as partial causes of later appropriate events, and yet we cannot claim that they influence their effects in the sense defined by Lewis. It seems that examples like these decisively falsify Lewis's latest attempt at a counterfactual theory of causation.

But they also pose a threat to earlier counterfactual analyses, since they admit a simple modification that introduces a particularly recalcitrant type of asymmetric redundancy. Thus, causation without influence but with redundancy seems to undermine the entire program of the counterfactual analysis of causation. Yet before we jump to conclusions let us recall that philosophers of science caution that negative evidence counts against a given theory only if there are viable alternatives to it that can accommodate this evidence.[18] And it looks like the main competitors of Lewis's approach encounter similarly serious difficulties when dealing with the cases of causality without influence (even without redundancy). To begin with, any general regularity approach will have the usual problems with defending the claim that the flipping of the switch implies the train's crash, as there may be numerous cases in which the railroad switch is changed and which are not followed by any crash due to the absence of the train. John Mackie's subtle version of the regularity approach that uses the notion of an INUS condition may deal with this problem, as it

[17] Yablo (2004) quotes Plato's distinction between causes and enabling conditions, i.e. "conditions that don't produce the effect themselves but create a context in which something else can do so" (p. 119). It looks like the firing of neuron C_1 satisfies Yablo's definition of *enablers*, as it is true that if D fired and C_1 did not, neuron E would not fire. Hence it may be claimed that C_1 is not a cause of E, but only enables D to cause E. However, Yablo later admits that enablers are full-fledged causes due to the existence of the counterfactual dependence between them and the effect, although they can be pragmatically counterindicated as such.

[18] See for instance (Thagard 1978).

requires only that a cause is a necessary part of a sufficient condition, but it suffers from the same setbacks as Lewis's theory due to cases of redundancy (Mackie 1965).

One of the most serious alternatives to the counterfactual approach to causation is the theory of causal processes based on the notion of the transfer of a conserved quantity. In Wesley Salmon's interpretation (accepted, with some modifications, by Phil Dowe), a process leading from c to e transmits a conserved quantity if it possesses the same amount of this quantity at all moments starting at c and ending at e, and in addition the process does not interact causally with any other process in a way that requires an exchange of that particular quantity.[19] For Salmon and Dowe, transmitting a conserved quantity in the above sense is the hallmark of causal processes that distinguishes them from so-called 'pseudo-processes'. But if we apply this definition to the railroad switch example we have to note that no conserved quantity is transferred from the pulling of the lever to the train's crash. The amount of energy, momentum, angular momentum that characterizes the movement of the lever is not equal to the amount of energy, momentum, etc. that is associated with the crash, nor is the latter in any way functionally dependent on the former. The transfer does occur, but it stops when the switch rails are moved into their new position, well before the arrival of the train.

Salmon-Dowe's theory of causation cannot be rescued by an appeal to the notion of causal interaction between processes. It may look like there are two interacting processes involved: the process leading from the movement of the lever to the new position of the switch rails, and the motion of the train. However, the definition of a causal interaction clearly demands that there be an exchange of a conserved quantity between two involved processes, and yet no such exchange is present in our case. The train neither loses nor gains any significant amount of energy or momentum upon passing the switch that would equal the gain or loss of energy/momentum of the process leading from the movement of the lever to the shifting of the switch rails. Hence the conserved quantity theory is clearly incapable of accounting for cases similar to the railroad switch case.[20]

Dowe admits that there are cases of seemingly causal processes that don't meet his requirement of the transmission of a conserved quantity, such as cases of causation by omission and prevention. His solution is to introduce the notion of

[19] See (Salmon 1997; Dowe 2000b, 2007).

[20] Salmon-Dowe's theory of causation is also threatened by a type of cause that Krajewski (1997) calls 'triggering causes'. A triggering cause is an event that initiates a release of previously accumulated energy. For instance, by removing a supporting rock we can cause a boulder to roll down the slope and to crash a tree in its path. However, the amount of energy/momentum spent for the removal of the obstacle stands in no relation to the amount of energy/momentum used for crashing the tree. A triggering cause releases potential energy that has been stored in an appropriate medium, and therefore is not connected by the relation of the transfer of a conserved quantity with the effect. It can be noted that with respect to the lack of a transfer of a conserved quantity triggering causes resemble causes used in the examples analyzed in this article. The only difference is that triggering causes are connected with their effects by the when-when dependence, for their occurrence is followed immediately (or after some fixed period of time) by the main process leading to the effect, whereas the time of the occurrence of a 'priming cause' stands in no fixed relation to the time of its effect. An example of a priming cause that is analogous to the aforementioned boulder example would be the action of removing an obstacle in the path of the boulder prior to an uncorrelated dislodging of the boulder.

pseudo-causation that could cover the cases that the conserved quantity theory cannot explain. He even formulates specific definitions of quasi-causation by omission and by prevention in terms of his 'genuine' causal processes and interactions.[21] It is very likely that a similar strategy could be applied to the cases of causation without influence. This suggests generally that we could always try to explain away obstinate cases that defy our favorite theory of causation by inventing for them a special category under the convenient umbrella of 'pseudo-causation'. Whether this would be the best way of rescuing Lewis's latest theory of causation from the quagmire of causation without influence remains to be seen. However, I think that another more promising plan of action may be available for the Lewisians.

Lewis's original theory of causation appears to be very well suited to deal with cases such as the railroad switch case without redundancy, as our intuitive judgment that they are indeed cases of causation is most likely based on the existence of the counterfactual dependence between the cause and the effect. Lewis replaced his stepwise counterfactual dependence theory with the influence theory in order to deal with the problem of preemption, and the abundant cases of causation without influence clearly prove that this was a step in the wrong direction. I believe that it may be prudent to give up the influence approach in favor of the good old theory, and to try again to tackle the real enemy, which is the problem of preemption (redundant causation).[22]

So far all efforts to solve the preemption problem have been directed toward finding one universal formula for the causal relation that would produce the right answer in each particular case of redundant asymmetric causation. Thus the standard is set very high. It may be noted that in analogous troublesome cases that are encountered in science we are very often satisfied with much less ambitious temporary solutions. In science it is customary to shield our favorite theories against single and atypical counterexamples with all sorts of 'idealizations' or *ceteris paribus* clauses. Why can't the same be done with philosophical theories? Many cases of redundant causation that philosophers are so eager to invent may be claimed to belong to the category of non-standard, marginal cases that our theory should not be required to explain directly. True, it would be ideal if we had a theory

[21] Incidentally, Lewis himself applied a similar strategy in order to deal with causation by omission, as it cannot be directly accommodated by his counterfactual theory due to the fact that omissions are (most probably) not events. Lewis suggested that the special case of causation by omission may be treated separately with the help of a special type of counterfactual (1986, pp. 189–193).

[22] Another strategy of how to deal with the problem of preemption, not mentioned in this article, has been briefly suggested by Field (2003, p. 452). The method recommended by Field is based on the modification of the requirement of counterfactual dependence between the cause and the effect to include cases of *conditional* counterfactual dependence. Field claims that it should be sufficient for causation to occur if the effect is counterfactually dependent on the cause given that we fix certain facts. The main problem is of course how to characterize generally the types of facts that are allowed to be fixed, without making an implicit reference to the notion of cause. Field illustrates his method with the case of early preemption, arguing that Suzy's throw counts as a cause, because if we hold fixed the fact that Billy didn't throw, the window's shattering does depend counterfactually on Suzy's throw. However, in the case of late preemption the situation is not so clear. We could argue either way that the shattering is counterfactually dependent on Billy's or Suzy's throw when we fix that the other throw did not occur. A more sophisticated version of the approach based on the conditional counterfactual dependence is presented in (Yablo 2004).

of causation that would be capable of delivering the correct analysis in every imaginable scenario and set-up, but sometimes we have to settle for less. A much more reasonable expectation in various cases of preemption is that we be able to produce a theoretical explanation of why our intuitive, pre-theoretical judgments depart from the verdict of our theory. Such an explanation may invoke for instance an apparent asymmetry between the preempted and preempting causes that is not taken into account by our theory due to the context-dependency of this asymmetry. If we were able to find in each known case of preemption a reasonable and principled basis of our judgment that selects one event as the preempting cause and another as the preempted one, then we should be satisfied with that sort of explanation even if this basis varied slightly in each case and thus was incapable of being incorporated into our general theory.[23]

Let me clarify these sketchy remarks a bit using the late preemption case as an example. As we have seen, neither Suzy's nor Billy's throw taken separately is counterfactually connected with the shattering of the bottle; however their disjunction is. So, according to the basic counterfactual analysis, only the whole disjunction of both events counts as a cause of the shattering. But now we should observe that the situation is not fully symmetric with respect to both disjuncts. This can be seen by comparing two closest possible worlds: one in which Suzy's throw is present but Billy's throw is absent, and the other one in which it is the other way around. Although in both worlds the required effect occurs, in the first world it is virtually identical with the actual shattering of the bottle, while in the second world the shattering differs in certain respects from the actual one due to the fact that in this particular world it is Billy's stone and not Suzy's which crashes the bottle. This asymmetry can be used to explain our inclination to single out one throw rather than the other as a cause of the shattering. Of all the contributing components of the disjunctive cause we tend to select the one whose absence would lead to the greatest modification of the effect, and this explains why we think it was Suzy's throw which caused the shattering.

But in other cases the asymmetry between the components of a disjunctive cause may have an altogether different basis. As we have seen, the potential causes in the case of causation without influence but with redundancy are symmetrical with respect to their power to modify the effect. In the neuronal network example discussed earlier, the firing of neuron F is counterfactually dependent on the disjunction of the firings of two priming neurons C_1 and C_2, but it is not dependent on any separate firing. Moreover, the firing of F would be qualitatively the same, even if the intermediary neuron E was primed by any of the two neurons C_1 or C_2 individually. In this case the asymmetry which can explain our preference of C_1 over C_2 is a temporal one. We pick the earlier of the two firings as a cause, because

[23] I believe that this suggestion goes in the same general direction as the proposal made by Hall (2004). Hall stresses that a theoretical account of causality does not have to recover all of our firm pre-theoretical intuitions. Rather, we should aim at construing the notion of causation in such a way that it can serve a useful theoretical purpose—for instance, helping us understand better the nomological structure of the world. According to Hall, cases of preemption teach us something important about causation—that it should be relatively stable, meaning that it should not go away when some extrinsic changes have been made.

we believe that the later firing did not make any significant change to the state of the already primed neuron E.

I suppose that similar asymmetry-based explanations can be provided in virtually all known cases of redundant causation, including trumping preemption. If this is correct, perhaps we should not worry extensively about the fact that the elementary counterfactual theory of causation does not immediately deliver the expected results in some of these cases. As long as the explanation of such a failure given in terms of the asymmetry between the components of the disjunctive cause is not blatantly ad hoc, this strategy of saving our best theory of causation from refutation should be seen as methodologically acceptable.

Acknowledgments I would like to express my gratitude to two anonymous referees for their extensive and stimulating comments which led to a substantial improvement of the paper.

References

Bigaj, T. (2005). Causes, conditions and counterfactuals. *Axiomathes, 15*, 599–619.

Coady, D. (2004). Preempting preemption. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 325–339). Cambridge, MA: The MIT Press.

Collins, J., Hall, N., & Paul, L. A. (2004). Counterfactuals and causation: History, problems and prospects. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals*. Cambridge, MA: The MIT Press.

Dowe, P. (2000a). *Is causation influence*? Unpublished manuscript.

Dowe, P. (2000b). *Physical causation*. New York: Cambridge University Press.

Dowe, P. (2007). Causal processes. In E. N. Zalta (Ed.). *The Stanford Encyclopedia of Philosophy (Fall 2007 edition)*. URL: http://plato.stanford.edu/archives/fall2007/entries/causation-process/.

Field, H. (2003). Causation in a physical world. In M. J. Loux & D. W. Zimmerman (Eds.), *The Oxford handbook of metaphysics* (pp. 435–460). Oxford: Oxford University Press.

Hall, N. (2000). Causation and the price of transitivity. *The Journal of Philosophy, 97*(4), 198–222.

Hall, H. (2004). Rescued from the rubbish bin: Lewis on causation. *Philosophy of Science, 71*, 1107–1114.

Krajewski, W. (1997). Energetic, informational, and triggering causes. *Erkenntnis, 47*, 193–202.

Kvart, I. (2001). Lewis's 'causation as influence'. *Australasian Journal of Philosophy, 79*(3), 409–421.

Lewis, D. (1973). Causation. *The Journal of Philosophy, 70*, 556–567.

Lewis, D. (1986). Postscripts to "Causation". In *Philosophical papers* (Vol. II, PP. 172–213). Oxford, NY: Oxford University Press.

Lewis, D. (2000). Causation as influence. *The Journal of Philosophy, 97*(4), 182–197. Reprinted in an extended form as (Lewis 2004).

Lewis, D. (2004). Causation as influence. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 75–106). Cambridge, MA: The MIT Press.

Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly, 2.4*, 245–255, 261–264.

McDermott, M. (1995). Redundant causation. *British Journal for the Philosophy of Science, 46*, 523–544.

Menzies, P. (2004). Difference-making in context. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 139–179). Cambridge, MA: The MIT Press.

Salmon, W. (1997). Causality and explanation: A reply to two critiques. *Philosophy of Science, 64*, 461–477.

Schaffer, J. (2001). Causation, influence, and effluence. *Analysis, 61.1*, 11–19.

Schwarz, W. (2009). *David Lewis: Metaphysik und Analyse*. Germany: Mentis.

Thagard, P. R. (1978). Why astrology is a pseudoscience. *Proceedings of Philosophy of Science Association, 1*, 223–224.

Tooley, M. (2003). Causation and supervenience. In M. J. Loux & D. W. Zimmerman (Eds.), *The Oxford Handbook of Metaphysics* (pp. 386–434). Oxford: Oxford University Press.

Yablo, S. (2004). Advertisement for a Sketch of an outline of a prototheory of causation. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 119–137). Cambridge, MA: The MIT Press.